

# ANALISIS SEGMENTASI PIUTANG MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING UNTUK OPTIMALISASI PENAGIHAN DI WANTI SALON

Alya Rachmadini Safitrie<sup>1</sup>; Agus Bahtiar<sup>2</sup>;

Program Studi Komputerisasi Akuntansi<sup>1</sup>  
Program Studi Sistem Informasi<sup>2</sup>

STMIK IKMI Cirebon

<https://ikmi.ac.id/page/18/?lang=de>

<sup>1</sup>alyar3264@gmail.com;

(\* Corresponding Author : alyar3264@gmail.com

Published : 30 Desember 2025

**Abstract**—Effective accounts receivable management plays a critical role in maintaining cash flow stability and business continuity, especially in service sectors such as beauty salons. This study aims to apply the K-Means Clustering algorithm to segment customers based on receivables data from Wanti Salon. The dataset used consists of 1000 customer entries with primary attributes being billing amount and receivable age in days. The research follows the Knowledge Discovery in Databases (KDD) framework, encompassing data selection, preprocessing, transformation, data mining, and evaluation stages. Experiments were conducted using RapidMiner with cluster configurations ranging from K=2 to K=5. The results demonstrate that the configuration with K=3 yields the most optimal segmentation by categorizing customers into three risk groups: low, medium, and high. Performance evaluation using the within centroid sum of squares supports the finding that K=3 provides the best balance between cluster compactness and business interpretability. This segmentation enables management to design more effective and adaptive collection strategies. Therefore, the implementation of clustering algorithms such as K-Means proves to be a data-driven solution for receivables management and can be replicated in other service sectors to support more strategic and analytically grounded decision-making.

**Keywords:** K-Means; Receivables; Customer Segmentation; Data Mining; RapidMiner

**Abstrak**—Manajemen piutang yang efektif merupakan faktor penting dalam menjaga stabilitas arus kas dan kelangsungan bisnis, khususnya pada sektor jasa seperti salon kecantikan. Penelitian ini bertujuan untuk menerapkan algoritma K-Means Clustering dalam melakukan segmentasi pelanggan berdasarkan data piutang dari Wanti Salon. Dataset yang digunakan terdiri dari 1000 entri pelanggan dengan atribut utama yaitu jumlah tagihan dan usia piutang dalam hari. Pendekatan Knowledge Discovery in Databases (KDD) digunakan sebagai kerangka kerja penelitian, yang mencakup tahapan seleksi data, pra-proses, transformasi, data mining, serta evaluasi. Eksperimen dilakukan dengan menggunakan perangkat lunak RapidMiner pada konfigurasi jumlah cluster K=2 hingga K=5. Hasil penelitian menunjukkan bahwa konfigurasi K=3 menghasilkan segmentasi paling optimal dengan mengelompokkan pelanggan ke dalam tiga kategori risiko: rendah, sedang, dan tinggi. Evaluasi performa berdasarkan nilai *within centroid sum of squares* juga mendukung bahwa nilai K=3 memberikan keseimbangan terbaik antara kekompakan cluster dan interpretabilitas bisnis. Segmentasi ini memberikan dasar yang kuat bagi manajemen dalam merancang strategi penagihan yang lebih efektif dan adaptif. Dengan demikian, penerapan algoritma clustering seperti K-Means dapat menjadi solusi berbasis data dalam pengelolaan piutang, serta dapat direplikasi pada sektor jasa lainnya untuk mendukung pengambilan keputusan yang lebih strategis dan berbasis analitik.

**Kata Kunci:** K-Means; Piutang Usaha; Segmentasi Pelanggan; Data Mining; RapidMiner

## INTRODUCTION

Dalam dunia bisnis, pengelolaan piutang usaha merupakan aspek yang sangat krusial dalam

menjaga keberlangsungan dan kesehatan finansial perusahaan. Piutang usaha, yang merujuk pada jumlah uang yang harus dibayarkan oleh pelanggan

atas jasa atau barang yang telah diterima, sering kali menjadi penyebab utama gangguan arus kas jika tidak dikelola dengan baik. Khususnya dalam sektor jasa seperti salon kecantikan, tantangan dalam pengelolaan piutang menjadi semakin kompleks karena frekuensi transaksi yang tinggi dan variasi perilaku pembayaran dari pelanggan. Oleh karena itu, diperlukan pendekatan analitik yang tepat untuk mengelompokkan pelanggan berdasarkan karakteristik piutang mereka agar strategi penagihan dapat dilakukan secara efisien dan tepat sasaran[1], [2].

Seiring dengan berkembangnya teknologi informasi, metode data mining seperti clustering menjadi sangat relevan untuk diterapkan dalam analisis data piutang. Clustering adalah salah satu teknik dalam unsupervised learning yang bertujuan untuk mengelompokkan data berdasarkan kemiripan fitur tanpa adanya label atau target yang ditentukan sebelumnya. Salah satu metode clustering yang populer dan banyak digunakan dalam analisis bisnis adalah K-Means Clustering. Metode ini bekerja dengan mengelompokkan data ke dalam K kelompok berdasarkan jarak Euclidean terdekat dari centroid tiap kelompok. Dengan pendekatan ini, perusahaan dapat memperoleh wawasan baru mengenai perilaku pelanggan, mengidentifikasi kelompok berisiko tinggi dalam pembayaran, serta merancang strategi penagihan yang lebih efektif[3], [4], [5].

Dalam konteks penelitian ini, analisis dilakukan terhadap data piutang pelanggan di Wanti Salon, sebuah usaha jasa kecantikan dengan jumlah pelanggan yang besar dan beragam. Dataset yang digunakan mencakup informasi tentang jumlah tagihan, usia piutang dalam hari, serta status pembayaran pelanggan. Berdasarkan atribut-atribut tersebut, dilakukan proses clustering dengan menggunakan algoritma K-Means pada berbagai konfigurasi jumlah cluster (K=2 hingga K=5). Tujuannya adalah untuk menemukan konfigurasi yang paling optimal dalam mengelompokkan pelanggan berdasarkan tingkat risiko pembayaran piutang[6], [7], [8].

Beberapa penelitian terdahulu telah membuktikan efektivitas metode K-Means Clustering dalam mengelompokkan data pelanggan berdasarkan berbagai variabel. Han et al. (2012) dalam bukunya "Data Mining: Concepts and Techniques" menjelaskan bahwa K-Means merupakan metode clustering yang efisien untuk data skala besar dan mampu memberikan segmentasi yang interpretatif untuk keperluan bisnis. Penelitian oleh Wibowo dan Susanto (2021) juga menunjukkan bahwa K-Means berhasil digunakan untuk segmentasi pelanggan pada bisnis ritel berdasarkan frekuensi transaksi dan nilai

pembelian, yang kemudian menjadi dasar dalam penyusunan strategi pemasaran yang lebih personal[9], [10].

Di bidang keuangan, penelitian oleh Prasetyo et al. (2020) memanfaatkan algoritma K-Means untuk mengelompokkan debitur berdasarkan kelayakan kredit dan tingkat risiko gagal bayar. Hasil penelitian tersebut menunjukkan bahwa segmentasi menggunakan K-Means mampu mengidentifikasi kelompok debitur dengan risiko tinggi yang memerlukan perhatian khusus dari manajemen keuangan. Hal ini menunjukkan bahwa pendekatan clustering tidak hanya bermanfaat dalam pemasaran tetapi juga dalam pengelolaan risiko keuangan [11], [12], [13].

Dalam konteks piutang usaha, clustering dapat digunakan untuk membedakan antara pelanggan yang sering membayar tepat waktu, yang sering menunda pembayaran, dan yang cenderung gagal bayar. Segmentasi ini sangat membantu dalam proses pengambilan keputusan, terutama dalam menentukan skema penagihan atau pemberian diskon kepada pelanggan tertentu. Dengan adanya pengelompokan berdasarkan karakteristik piutang, perusahaan dapat lebih fokus dalam menetapkan prioritas penagihan dan mengurangi risiko piutang tak tertagih.[14], [15]

Permasalahan utama yang dihadapi dalam penelitian ini adalah bagaimana mengelompokkan pelanggan berdasarkan karakteristik piutang mereka secara efektif agar strategi penagihan dapat dilakukan dengan lebih efisien. Saat ini, banyak perusahaan jasa termasuk salon kecantikan yang masih menggunakan pendekatan konvensional dalam menangani piutang, yaitu dengan memperlakukan semua pelanggan secara seragam tanpa mempertimbangkan tingkat risiko pembayaran. Pendekatan ini kurang efektif karena tidak memberikan informasi yang cukup bagi manajemen untuk memprioritaskan usaha penagihan.

Selain itu, kurangnya pemanfaatan data historis pelanggan juga menjadi masalah serius. Data transaksi dan pembayaran yang tersedia sering kali tidak dimanfaatkan secara optimal untuk analisis yang lebih mendalam. Hal ini menyebabkan potensi kerugian akibat piutang tak tertagih menjadi tinggi. Oleh karena itu, dibutuhkan metode analitik yang mampu mengolah data historis tersebut menjadi informasi yang berguna dalam pengambilan keputusan strategis.

Untuk menjawab permasalahan tersebut, penelitian ini mengusulkan penggunaan algoritma K-Means Clustering sebagai solusi segmentasi pelanggan berdasarkan data piutang. Dengan mengelompokkan pelanggan ke dalam beberapa kategori risiko, perusahaan dapat menyusun

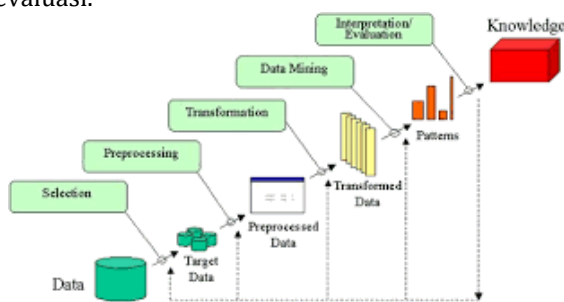
strategi penagihan yang berbeda-beda untuk tiap kelompok. Misalnya, pelanggan dengan piutang tinggi dan usia piutang yang lama dapat diprioritaskan untuk penagihan intensif, sedangkan pelanggan dengan piutang rendah dan usia piutang pendek mungkin cukup dipantau secara rutin[16], [17], [18].

Dalam pelaksanaan penelitian, digunakan perangkat lunak RapidMiner untuk menerapkan algoritma K-Means terhadap dataset yang berisi informasi jumlah tagihan, usia piutang, dan status bayar dari 1000 pelanggan. Model dievaluasi pada beberapa nilai K untuk menentukan jumlah cluster yang paling tepat. Setiap hasil clustering kemudian dianalisis secara visual dan deskriptif untuk mengetahui karakteristik masing-masing cluster. Dengan cara ini, manajemen dapat lebih mudah memahami profil pelanggan dan mengambil tindakan yang tepat.

Penelitian ini diharapkan dapat memberikan kontribusi nyata dalam penerapan teknik data mining untuk pengelolaan piutang yang lebih efisien, terutama di sektor jasa seperti salon kecantikan yang sering menghadapi tantangan serupa. Selain itu, pendekatan ini juga dapat diadopsi oleh perusahaan jasa lainnya sebagai bagian dari sistem manajemen keuangan yang lebih cerdas dan berbasis data.

## MATERIALS AND METHODS

Penelitian ini mengacu pada kerangka kerja Knowledge Discovery in Databases (KDD), yang terdiri dari lima tahap utama: seleksi data, praproses data, transformasi, data mining, dan evaluasi.



Gambar 1. KDD

Seleksi DataData yang digunakan dalam penelitian ini berasal dari sistem pencatatan transaksi di Wanti Salon. Dataset terdiri dari 1000 entri pelanggan dengan atribut utama: ID\_Pelanggan, Jumlah\_Tagihan (dalam Rupiah), Usia\_Piutang\_Hari (usia piutang sejak tanggal jatuh tempo), dan Status\_Bayar (Lunas atau Belum Lunas). Hanya data dengan atribut lengkap dan relevan yang diseleksi untuk proses selanjutnya.

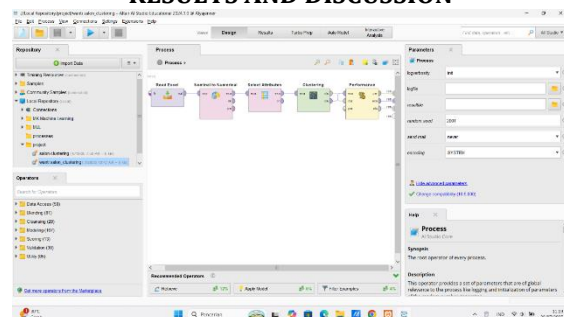
Praproses DataTahap ini mencakup pembersihan data dari entri yang tidak valid, duplikat, dan noise. Kolom Status\_Bayar dikonversi menjadi informasi kategorikal terpisah untuk keperluan validasi dan analisis deskriptif. Data numerik seperti Jumlah\_Tagihan dan Usia\_Piutang\_Hari dinormalisasi agar setiap atribut memiliki bobot yang setara dalam proses clustering.

Transformasi DataTransformasi dilakukan untuk mengonversi data ke dalam format yang sesuai untuk algoritma K-Means. Dua atribut numerik yaitu Jumlah\_Tagihan dan Usia\_Piutang\_Hari dipertahankan sebagai variabel input, karena keduanya mencerminkan potensi risiko piutang dari pelanggan.

Data MiningAlgoritma K-Means Clustering diterapkan menggunakan RapidMiner. Eksperimen dilakukan pada beberapa konfigurasi nilai K yaitu 2, 3, 4, dan 5. Hasil dari tiap nilai K dianalisis untuk melihat distribusi dan karakteristik masing-masing cluster. Penentuan nilai K terbaik dilakukan dengan mempertimbangkan interpretabilitas hasil dan visualisasi plot.

Evaluasi dan InterpretasiEvaluasi dilakukan secara deskriptif melalui analisis nilai rata-rata setiap atribut pada masing-masing cluster, serta interpretasi visual menggunakan scatter plot. Diperhatikan pula distribusi status bayar dalam masing-masing cluster untuk mengidentifikasi korelasi antara kelompok risiko dan kemungkinan keterlambatan pembayaran. Hasil analisis ini digunakan untuk menyarankan strategi penagihan yang tepat bagi tiap segmen pelanggan.

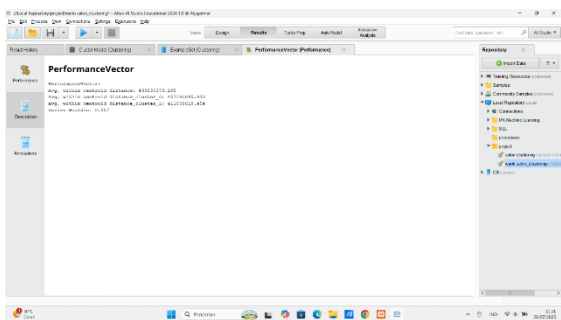
## RESULTS AND DISCUSSION



Gambar 2. Desain

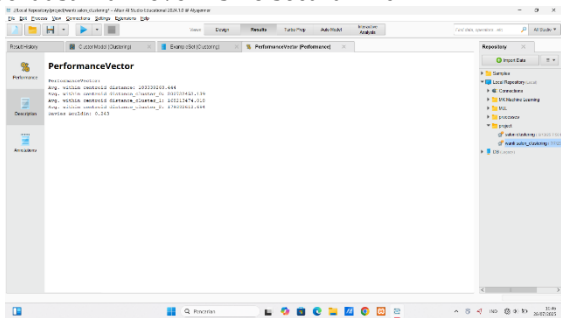
Gambar 2. pertama menampilkan desain proses kerja dalam RapidMiner yang digunakan untuk melakukan segmentasi pelanggan berdasarkan data piutang. Proses diawali dengan operator Read Excel yang berfungsi untuk mengimpor dataset pelanggan dari file Excel. Kemudian dilanjutkan dengan Select Attributes untuk memilih atribut yang relevan, yaitu Jumlah\_Tagihan dan Usia\_Piutang\_Hari, yang menjadi dasar dalam analisis clustering. Operator Normalize digunakan untuk menyeimbangkan

skala antar fitur numerik agar algoritma K-Means tidak bias terhadap nilai absolut. Selanjutnya, operator Clustering (K-Means) digunakan untuk mengelompokkan data, dan hasilnya dievaluasi dengan operator Performance guna menilai kualitas model. Proses ini menunjukkan penerapan pipeline yang sistematis dan sesuai dengan alur KDD (Knowledge Discovery in Databases).



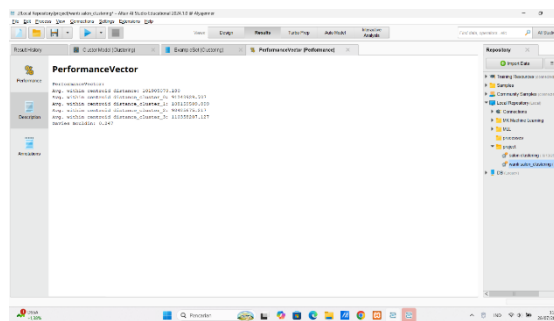
Gambar 3. Hasil Plot K2

Gambar 3. menunjukkan hasil evaluasi performa clustering dengan jumlah cluster (K) = 2. Nilai within centroid sum of squares tercatat sebesar 439229117.823. Nilai ini menggambarkan total jarak dari setiap titik data ke centroid cluster-nya masing-masing. Semakin kecil nilai ini, semakin baik pemodelan dalam hal kekompakan antar data dalam setiap cluster. Namun, dengan hanya dua cluster, segmentasi masih tergolong terlalu umum dan kurang mampu membedakan pelanggan berdasarkan level risiko secara rinci.



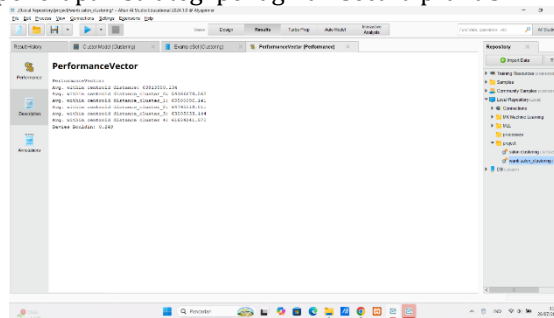
Gambar 4. Hasil Plot K3

Pada gambar 4, ditampilkan hasil evaluasi saat jumlah cluster ditetapkan sebanyak 3 (K=3). Nilai within centroid sum of squares menurun menjadi 273315050.14, menandakan peningkatan kekompakan kelompok dibandingkan konfigurasi K=2. Nilai ini menunjukkan bahwa penambahan jumlah cluster mampu memecah kelompok besar menjadi kelompok yang lebih homogen. Dalam konteks bisnis, konfigurasi ini sangat relevan karena mampu membagi pelanggan ke dalam kategori risiko rendah, sedang, dan tinggi secara lebih akurat.



Gambar 5. Hasil Plot K4

Gambar 5. menampilkan performa model dengan K=4, menghasilkan nilai *within centroid sum of squares* sebesar 214130153.419. Nilai ini lebih kecil dari K=3, yang secara statistik menunjukkan peningkatan kekompakan cluster. Namun, secara interpretatif, pemisahan menjadi empat kelompok justru dapat menimbulkan ambiguitas. Beberapa cluster mungkin memiliki karakteristik yang tumpang tindih, sehingga menyulitkan dalam penerapan strategi penagihan secara praktis.



Gambar 6 Hasil Plot K5

Gambar 5. menggambarkan hasil eksperimen pada K=5. Nilai *within centroid sum of squares* turun kembali menjadi 170693508.10, mencerminkan distribusi yang makin granular. Namun, terlalu banyak cluster justru membuat segmentasi menjadi terlalu spesifik dan kurang bermakna secara bisnis. Terdapat risiko overfitting dan ketidakefisienan dalam pengambilan keputusan, karena beberapa cluster hanya berisi sedikit anggota yang tidak signifikan secara operasional.

Dari hasil pengujian pada berbagai nilai K, dapat disimpulkan bahwa nilai K=3 merupakan konfigurasi terbaik. Segmentasi yang dihasilkan mampu memetakan pelanggan ke dalam kelompok risiko rendah, sedang, dan tinggi secara jelas dan mudah dipahami. Hal ini memungkinkan manajemen Wanti Salon untuk menyusun strategi penagihan yang lebih efektif, seperti memprioritaskan penagihan untuk pelanggan di cluster risiko tinggi dan memberikan insentif atau pengingat ringan untuk pelanggan di cluster risiko rendah.

Dengan visualisasi scatter plot dan analisis nilai centroid masing-masing cluster, manajemen dapat mengembangkan sistem manajemen piutang berbasis risiko yang lebih adaptif dan responsif terhadap perilaku pelanggan. Pendekatan ini terbukti efektif dalam mendukung keputusan bisnis yang berbasis data dan meningkatkan potensi keberhasilan dalam proses penagihan.

### CONCLUSION

Penelitian ini menunjukkan bahwa penerapan algoritma K-Means Clustering dapat memberikan solusi efektif dalam mengelompokkan pelanggan berdasarkan karakteristik piutang. Dengan menggunakan data jumlah tagihan dan usia piutang, segmentasi pelanggan menjadi lebih informatif dan dapat digunakan untuk menentukan strategi penagihan yang tepat sasaran. Hasil eksperimen pada konfigurasi K=2 hingga K=5 menunjukkan bahwa nilai K=3 merupakan segmentasi paling optimal, karena berhasil membagi pelanggan ke dalam tiga kategori risiko: rendah, sedang, dan tinggi. Temuan ini tidak hanya relevan bagi Wanti Salon, tetapi juga dapat menjadi referensi bagi perusahaan jasa lainnya dalam mengelola piutang secara proaktif dan berbasis data

### REFERENCE

- [1] M. H. Abdurrahman, E. Suhartono, and E. Wulandari, "DETEKSI KUALITAS KEMURNIAN SUSU SAPI MELALUI PENGOLAHAN CITRA DIGITAL MENGGUNAKAN METODE SCALE INVARIANT FEATURE TRANSFORM (SIFT) DAN KLASIFIKASI K-NEAREST NEIGHBOR (KNN) QUALITY DETECTION OF COW'S MILK PURITY USING DIGITAL IMAGE PROCESSING METHOD OF SCALE INVARIANT FEATURE TRANSFORM WITH K-NEAREST NEIGHBOR CLASSIFICATION."
- [2] P. Pengolahan Citra Klasifikasi Jenis Mangga Menggunakan Metode K-Nearst Neighbor, "Pemodelan Pengolahan Citra Klasifikasi Jenis Mangga Menggunakan Metode K-Nearst Neighbor," *Jurnal Sains dan Teknologi*, vol. 5, no. 1, p. |pp, 2023, doi: 10.55338/saintek.v5i1.1422.
- [3] R. Widadi and D. Zulherman, "Klasifikasi Pergerakan Tangan dan Kaki Berbasis Sinyal EEG Menggunakan Common Spatial Patterns dan Multilayer Perceptron Backpropagation," *Jurnal Telematika*, vol. 14, no. 2, 2020.
- [4] P. Antiperovitch *et al.*, "Continuous Atrial Fibrillation Monitoring From Photoplethysmography: Comparison Between Supervised Deep Learning and Heuristic Signal Processing," *JACC Clin Electrophysiol*, vol. 10, no. 2, pp. 334–345, Feb. 2024, doi: 10.1016/j.jacep.2024.01.008.
- [5] T. M. T. A. Hamid, R. Sallehuddin, Z. M. Yunos, and A. Ali, "Ensemble filters with harmonize PSO-SVM algorithm for optimal hearing disorder prediction," *Neural Comput Appl*, vol. 35, no. 14, pp. 10473–10496, May 2023, doi: 10.1007/s00521-023-08244-2.
- [6] E. Agustin, A. Eviyanti, and N. Lutvi Azizah, "Deteksi Penyakit Epilepsi Melalui Sinyal EEG Menggunakan Metode DWT dan Extreme Gradient Boosting," *Jurnal Media Informatika Budidarma*, vol. 7, no. 1, pp. 117–127, 2023, doi: 10.30865/mib.v7i1.5412.
- [7] N. Elsayed, Z. Saad Zaghloul, and M. Bayoumi, "Brain Computer Interface: EEG Signal Preprocessing Issues and Solutions," 2017.
- [8] D. Candra and R. Novitasari, "Klasifikasi Sinyal EEG Menggunakan Metode Fuzzy C-Means Clustering (FCM) Dan Adaptive Neighborhood Modified Backpropagation (ANMBP)," *Jurnal Online Fakultas Sains dan Teknologi*, 2022, [Online]. Available: [http://sccn.ucsd.edu/~arno/fam2data/publicly\\_](http://sccn.ucsd.edu/~arno/fam2data/publicly_)
- [9] A. Fitriawati Zakiyyah, "Klasifikasi Emosi Untuk Mengetahui Pengalaman Emosional Melalui Sinyal EEG Menggunakan Algoritma Artificial Neural Network," *Jurnal Sains Aplikasi Komputasi dan Teknologi*, vol. 3, no. 2, p. 40, 2021.
- [10] R. Widadi, B. Arif Widodo, and D. Zulherman, "Klasifikasi Sinyal EEG pada Sistem BCI Pergerakan Jari Manusia Menggunakan Convolutional Neural Network," *Jurnal Techno.com*, vol. 19, no. 4, pp. 459–467, 2020.
- [11] A. Purnomo and H. Tjandrasa, "IMPROVED DEEP LEARNING ARCHITECTURE WITH BATCH NORMALIZATION FOR EEG SIGNAL PROCESSING," *JUTI: Jurnal Ilmiah Teknologi Informasi*, vol. 19, no. 1, p. 19, Jan. 2021, doi: 10.12962/j24068535.v19i1.a1023.
- [12] T. M. M. Keumala, M. Melinda, and S. Syahrial, "Decision tree method to classify the electroencephalography-based emotion data," *JURNAL INFOTEL*, vol. 14, no. 1, pp. 37–49, Feb. 2022, doi: 10.20895/infotel.v14i1.750.
- [13] A. Norma Utami Jurusan Matematika and F. Matematika dan Ilmu Pengetahuan Alam,

- "Klasifikasi Gangguan Tidur REM Behaviour Disorder Berdasarkan Sinyal EEG Menggunakan Machine Learning," *Jurnal Sistem Cerdas*, 2020.
- [14] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Comput Biol Med*, vol. 100, pp. 270–278, Sep. 2018, doi: 10.1016/j.compbimed.2017.09.017.
- [15] A. Eviyanti, H. Hindarto, Sumarno, and H. A. A. Duddin, "Epilepsi detection system based on EEG record using neural network backpropagation method," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Nov. 2019. doi: 10.1088/1742-6596/1381/1/012037.
- [16] A. Novianto and M. D. Anasanti, "Autism Spectrum Disorder (ASD) Identification Using Feature-Based Machine Learning Classification Model," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 17, no. 3, p. 259, Jul. 2023, doi: 10.22146/ijccs.83585.
- [17] N. Nurzilla, "PREDIKSI PERTUMBUHAN TUMOR KANKER PAYUDARA MENGGUNAKAN MODEL REGRESI LINEAR BERBASIS MACHINE LEARNING," 2024.
- [18] M. Desai and M. Shah, "An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN)," Jan. 01, 2021, *KeAi Communications Co.* doi: 10.1016/j.ceh.2020.11.002.